

20 MAR 1980

MEMORANDUM FOR: Deputy Director for Administration
STATINTL FROM: [REDACTED] M.D.
Director of Medical Services
SUBJECT: IG Draft Report - Chapter IX Professional
Applicant Test Battery (PATB)

1. Attached herewith is the response to Chapter IX of the IG Draft Report on Recruitment. This review has been prepared by three Agency staff Ph.D. psychologists, in consultation with three highly qualified consultants, two of whom are nationally recognized experts. They are identified by name on the title page of the report.

2. I fully endorse and support their comments, conclusions, and recommendations. It is my fervent hope that this document will go forward, along with the IG Report, to the DDCI and DCI.

STATINTL



Att (4 copies)

STATINTL

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130003-5

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130003-5

Introduction

This paper responds to the IG report on the use of PATB in selecting personnel for Agency employment. The IG investigation is contained in four papers: (1) a draft of text summarizing the investigation which, we understand, will be included as a chapter in an overall IG report on the Agency's entire personnel recruitment, processing, and selection system; (2) a summary report on the PATB prepared by two IG external consultants; (3) an appendix prepared by the two consultants in which PATB validity and reliability data are reviewed; and (4) a second appendix, also prepared by the two consultants, in which narrative reporting of PATB results by PSS psychologists is discussed.

In this paper, only the first of these four reports is addressed. We feel, however, that the other three reports also require point-by-point commentary, and we are prepared to provide this at a later time.

Major Findings and Conclusions

Overall, the IG inspection of PATB must be regarded as defective, and its recommendations misguided. We find the IG report on PATB seriously in error in several major respects. The review of PATB by the IG consultants was based on misunderstanding and incorrect assumptions. Their

judgments about PATB are based on the implicit assumptions
that:

- (1) the PATB is used, or intended to be used, as a primary screening device for all employees, and should be so validated.
- (2) the Agency is committed to using PATB in a line capacity to review and make decisions, or at least recommendations, on every applicant.
- (3) PSS has sufficient resources for ongoing development and refinement of the test battery and has continuing access to the performance appraisals needed to conduct validation research.
- (4) the Agency has agreed to provide the level of support needed to make all of this possible.

These errors are predictable when outside consultants are drawn from an academic rather than an industrial setting, where the major competence in these areas resides today, and when the primary objects used for study are internal reports written for reading by non-psychological users, not for academic review. Unhappily, there is substantial evidence, fully documented in the succeeding paragraphs, that a professional review, properly objective and properly attentive to the organizational realities within CIA, was not conducted, and that the investigation is previously marred by errors, omissions, contradictions, and, on occasion, fantasy.

The report recommends major changes in the personnel management policies of CIA, concealed as criticisms of PATB

and PSS. We do not support the recommendations, for they would greatly increase the workload of PSS and the costs of testing without materially improving the selection process. The current policies that place PSS in a staff relation with Agency managers work well, and permit PSS to use its limited resources where they will do the most good.

Selection of the Consultants

The inspection report claims on page 1 that evidence concerning the reliability and validity of the PATB was examined by "...two nationally recognized experts in psychological testing." The selection of these consultants merits comment. The principal consultant, [REDACTED] was STATINTL selected by the IG staff from a sparse list of five names. How this list was generated is unknown to us, but its extreme brevity suggests a disturbing lack of familiarity with the field of psychological testing and measurement on the part of the IG staff.

A brief description of these two "nationally recognized experts" is in order at this point. [REDACTED] is currently 69 years of age and apparently has been retired from full-time employment since 1976. He received a Ph.D. degree from Columbia University in 1935 and appears to have spent the bulk, if not all, of his professional career at Columbia University Teachers College as a professor in psychology and education. [REDACTED] who also received a Ph.D. from Columbia, is 64 years of age and is currently a

professor of psychology and education at Columbia University Teachers College. She too appears to have spent the bulk, if not all, of her professional career in an academic setting. The background of the two consultants is noteworthy in that it helps explain the academic approach taken by them to the inspection of the PATB. Neither of these consultants has had; as far as we can determine, any direct experience in the operational day-to-day conduct and management of an applied testing program in an organizational setting. For purposes of inspecting an applied testing program, the designation of [REDACTED] as "nationally recognized experts" is inaccurate. A choice of a psychologist working in industry would have been far preferable, for such persons face the same problems CIA faces in the use of tests in employment selection.

Role of PATB in Personnel Selection

On pages 1 through 4 of their report, the inspectors challenge CIA policies about the "role of PATB in personnel selection" in the Agency. They begin on page 2 by implying that something is wrong with the fact that some components use the PATB whereas others do not. At present, PATB testing is a service which is available to those managers who find it useful in the evaluation and selection of applicants for professional positions in their components. PATB reports are advisory in nature and describe applicants in terms which may be of value to managers faced with selection decisions.

The PATB report is one of many inputs available to the manager and is in no way binding on him. Other inputs include academic transcripts, academic degrees, impressions formed during interviews, letters of recommendation, etc., all of which are weighted by the manager in making his selection decisions. The IG report contains no data of any kind to support their implied recommendation that these policies and procedures should be changed to avoid "inconsistencies." Nor does it contain data showing how the PATB report is weighted in relation to other selection inputs. The IG survey did collect some superficial statistics on how many managers find PATB reports useful and how many applicants report having taken PATB, but again we must stress that the IG investigation did not accomplish its stated objective to determine the "role of PATB in personnel selection." So-called "inconsistencies" in the use of PATB are described as "serious...on grounds of fairness [and] legality" and "unwise in terms of Agency goals in hiring the most qualified personnel," but these vague admonitions are not substantiated by data.

We do not support the inspectors' statement that "...a systematic policy on personnel selection practices and the role of PATB in personnel selection" is needed. Rather, we believe that the hiring of personnel should be left in the hands of individual managers, who know more about the unique requirements of the jobs they must fill than anyone else. We

further reject, in the absence of any empirical data that misuse of PATB has occurred, the notion that the basis of selection must be centralized. We do not favor policies that impose a universal requirement that tests be used as part of the selection process. That the IG report would make such recommendations in the absence of empirical data reflects, in our judgment, inadequate study of the problem. Their claim that the absence of such policies has "...a high potential for violating U.S. Equal Employment Opportunity Commission (EEOC) Guidelines on disparate treatment of applicants" seems to us deliberately misleading. There is no requirement in the guidelines that applicants for different and highly disparate professional jobs be evaluated by the same criteria. It is only when applicants for the same job are treated differently that potential for violation of the guidelines exists. Considering the large number of highly disparate professional jobs in the Agency, we consider it entirely proper that individual managers determine the criteria to be used in the selection of applicants to fill their professional positions. The IG report has presented no empirical data of any kind to justify the centralization of selection processes, and the inspectors' attempt to justify their argument for this on the basis of potential EEOC violation is based both on faulty reasoning and on misunderstanding of what the guidelines actually say.

PATB Reliability

On pages 4 and 5, the inspectors claim that overall PATB reliability is inadequate. They describe PATB reliability data as "fragmentary," and then state that "...since 1958 only one reliability study has been done..." on PATB. This latter statement is incorrect. The author of the single reliability study cited by the inspectors produced four other studies on PATB reliability in the same year as the one cited. The inspectors do not explain why they disregarded the other four studies conducted by this individual. The inspectors further argue that "...PATB reliability data are particularly inadequate for females and minorities."

Usually, when a test battery is developed, validated, and normed for use in personnel selection, a concerted effort is made to establish the reliability of the tests comprising the battery. This was done for PATB. Once the reliability of the tests is established, and the test battery put into use, there is usually little reason to question the reliability further. Hence, additional reliability studies are generally not done, unless some special reason arises to question further the reliability of tests comprising the battery. Thus, the inspectors' criticisms of reliability seem to us to be petty and unwarranted.

✓ However, the simplest way to resolve the question of whether or not the PATB has acceptable reliability is to

once again examine empirical data on this point. Hence, we collected data on a sample of applicants for professional employment in CIA who were PATB tested in 1978 and 1979. In order to address the criticisms of the inspectors of inadequate reliability data for minorities and females, we selected a sample in which both were generously represented. Since we have had very little time to conduct this exercise, we present here reliability data on a relatively small applicant sample for only seven of the PATB tests, but we are fully prepared to undertake a more comprehensive study at a later date. Presented in Table 1 are Spearman-Brown reliability estimates based on split-half score comparisons for the seven PATB tests. Reliability is computed separately for: (a) the entire sample of 426 applicants; (b) the 235 males and 191 females in the total sample; (c) the 228 whites and 198 blacks in the total sample; and (d) the 116 white males, 112 white females, 119 black males, and 79 black females comprising the total sample. Overall, the reliabilities presented in Table 1 are most impressive. For the total sample, all reliabilities exceed .85 and three exceed .90. The reliabilities computed separately for males and females and for whites and blacks also exceed .80 in every instance and several exceed .90 as well. It is not until the total sample is sub-divided on the basis of both sex and race, where the sub-samples become small, that any of the reliability estimates fall

TABLE 1
Spearman-Brown Reliabilities of Seven PATB Tests

PATB Tests	426 Applicants	235 Males	191 Females	228 Whites	198 Blacks	116 White Males	112 White Females	119 Black Males	79 Black Females
Vocabulary Mastery	.942	.940	.946	.924	.921	.930	.919	.922	.921
Reading Comprehension	.879	.865	.896	.824	.813	.839	.801	.776	.870
Abstract Reasoning	.871	.874	.861	.846	.829	.852	.823	.825	.818
Arithmetic Problems	.906	.915	.886	.884	.829	.904	.838	.818	.796
Foreign Language Aptitude	.964	.953	.971	.966	.896	.960	.965	.889	.906
Knowledge of World Events	.889	.882	.880	.860	.862	.877	.813	.840	.826
Interpretation of Data	.875	.890	.852	.836	.814	.869	.788	.820	.792

below .80, but it should be noted that reliabilities less than .80 are very close to, and in fact statistically indistinguishable from this value. The data presented in Table 1 demonstrate clearly that PATB has more than adequate reliability. Furthermore, the evidence shows that the PATB is reliable both for females and for minorities. The inspectors concluded, based on ^a_A superficial and incomplete study, that evidence for the reliability of PATB is not "encouraging." But the evidence we present here demonstrates that the PATB is reliable; these data refute the unsupported claims of the inspectors to the contrary.

PATB Validity

In Section C on pages 5, 6, and 7 of their report, the inspectors present a discussion of the validity of the PATB in which material facts are distorted and misrepresented. The work of the consultants was not helpful and obviously produced confusion in the minds of the inspectors. We hope our comments are helpful in sorting out several complex issues. Before responding to specific statements made in this section of their report, we wish to comment in general terms on the approach taken by the inspectors and their consultants to the examination of PATB validity. Following this, we shall respond to specific statements made by the inspectors in their draft chapter.

(1) IG test validation model. The inspectors state on page 6 that the evidence for the validity of PATB is "...fragmentary, very weak and unconvincing..." They do not, however, mention the difficulties in doing validation studies in an organizational setting, which are well described by their own consultants. They leave the reader with the impression that no efforts have been made to conduct proper validity research, and further imply that PSS should be charged with neglect. This overall conclusion on their part emanates not only from their inadequate review of the evidence for validity but also from their apparent efforts to generalize from the academic approach to test validation taken by their consultants. The rigid constraints imposed by a textbook approach to test validation preclude the utilization of this approach in any applied organizational setting unless that organization is willing to adopt, for an extended period, a random selection procedure

in the hiring of applicants to fill its vacant positions. Basically, determining with accuracy the true validity of an employment test requires that applicants for all positions of a given class be evaluated at the time of application for employment (by tests, interviews and other screening devices), but that no selection decisions be made on the basis of the information derived from any of the screening devices. Random selection is used at this point to fill position vacancies. Each such randomly selected employee must work on the job long enough to yield a reliable estimate of job performance. The information obtained from the tests and other screening techniques is then evaluated for its effectiveness in predicting job performance, usually through correlational analysis. In order to insure that the observed validation correlations are not unique to the particular individuals included in the initial random selection group, the entire procedure is repeated with another group

of applicants selected randomly. This step is called "cross-validation." If the correlations between the screening data and subsequent job performance hold up, then those screening techniques are said to have criterion-related validity and can then be used in the selection of future employees. It must be emphasized here that this general procedure must be applied to all selection techniques, not just tests, in order to establish their validity for personnel selection decisions.

It should be patently obvious, however, that a random selection model cannot be used to validate potential selection techniques in our organization. Employment of marginal and unsatisfactory applicants solely for the purpose of validating selection techniques would exact high cost in terms of impact on the ability of an intelligence organization to carry out its basic mission. Like it or not, the random selection model, which works very well in the pedagogical textbooks and on the classroom blackboards of academic theoreticians, has very little utility

in the real world, especially in an organizational setting such as an intelligence agency where success or failure is critically dependent upon selection of a cadre of professional employees of the highest possible caliber. Evaluation of the validity of PATB within the framework of a random selection model is clearly inappropriate. The inspectors may simply have failed to recognize this, but considering the strident manner in which their report is written, we believe theirs was a conscious decision to try to cast PATB in the worst possible light.

(2) Consequences of the random selection model.

As a result of the fact that random selection has never been and is not likely to be employed in the selection of Agency personnel, the ability of tests and other selection devices to show strong positive correlations with actual job performance is severely limited.

This phenomenon, called restriction of range, is caused by the lack of opportunity for tests and other selection instruments to predict poor performance since those likely to perform poorly are not hired to start with.

The demonstration of low or modest correlations between tests and job-related performance measures in an organization where only very few, and presumably the best, individuals are hired from the general applicant pool suggests that substantially larger validation correlations would be found had random selection been employed in the initial selection of applicants. Thus, whereas the low and modest correlations obtained in many PSS validation studies are interpreted by the inspectors as "...very weak and unconvincing..." such a pattern of validation coefficients is exactly what one would expect in an organizational setting where (a) the percentage of applicants selected for employment (i.e., the selection ratio) is low to start with, and (b) use of a random selection strategy is clearly inappropriate.

Again, we find it difficult to see how the inspectors could have failed to learn this from their consultants, and as a consequence we are left with no choice except to view this omission on their part as evidence either of bias or superficiality.

(3) The criterion problem. The most serious omission in the IG report on PATE validity is the failure of the inspectors and their consultants to discuss the performance criterion problem and its implications for test validation research. In its most basic form, test validation involves a search for significant correlations between test scores and measures of job performance. If these are found, the tests are said to have criterion-related validity and can then legitimately be used to select future applicants for those job settings. Tests, on the one hand, measure relatively unitary attributes. In fact, they are deliberately designed to do this. Individual tests, or scales within tests, are said to be factorially simple, i.e., they measure one, or at most a small number of very specific attributes. In measuring job performance, on the other hand, an attempt is made to devise a numerically simple measurement scale which accurately reflects the overall quality of an individual's work performance. Although the measurement scale used for this purpose

may appear to be simple, what it is being used to measure, viz., job performance, is very complex. Job-performance ratings are influenced by a myriad of factors, some of which are unique to a particular job, others common to perhaps a family of jobs. Measures of job performance, which may appear to be quite simple, are actually factorially very complex. Thus, when an attempt is made to correlate scores on a particular test with measures of job performance, that particular test will correlate with job performance only to the extent to which the factor (or factors) measured by the test is (are) represented by the performance measure. Hence, in most attempts to correlate tests with performance, one expects to find many tests correlating to a low or modest degree with the performance criterion measure rather than to find a few tests correlating to a high degree with the performance criterion. This is exactly the pattern of results obtained in most of the PSS test validation studies, and is totally consistent with the decision

made initially to develop a comprehensive, broad-based battery of tests for use as an input to Agency personnel selection. Use of a battery of tests as comprehensive as the PATB practically guarantees factorial overlap between PATB scores and measures of performance in Agency job settings. Factorial overlap in turn guarantees validity. But again, because tests are factorially simple and job-performance measures factorially complex, we expect to see in test validation studies many tests correlating a little rather than a few tests correlating a lot, which is exactly what we have found in most of our validation research. One must also take into consideration the fact that the small validation coefficients expected because of the criterion problem are further attenuated by the restriction of range resulting from the inability to use a random selection model (see above). The fact that PSS test validation studies have been able in almost every instance to demonstrate statistically significant correlations

with performance criterion measures is convincing testimony that PATB tests are very robust and sufficiently valid to justify their use in Agency personnel decision making. The failure of the IG consultants to consider these factors in their review of PSS validation research leaves the reader with the impression that if all validity coefficients are not high, something is wrong. We invite both the inspectors and their consultants to examine the book by Miner and Miner entitled Employee Selection Within the Law, which discusses in detail the impact on test validation of both restriction of range and the criterion problem.

- (4) The claim is made on page 6 of the IG report that the original selection of PATB tests and scales was not based on comprehensive analysis of jobs. This assertion is historically incorrect, and is directly contradicted by background information on the development of PATB provided to both the inspectors and their consultants. It would have been one thing for the inspectors to state that they regarded the

original analysis of jobs as inadequate by today's standards, but the inspectors do not mention that they were provided with the PSS Test Data Book, which reports much of the research data on which the construction of PATB was originally based, and which contains test-score profiles on a large number of Agency job groups. The original selection of sub-tests for PATB was based on detailed knowledge and understanding of jobs derived from years of extensive discussions with supervisors, managers, and employees of the basic requirements of a large number of Agency jobs. Prior to the development of the PATB, the selection of persons for Agency jobs was based in part on comprehensive intensive assessment of applicants for those jobs and subsequent monitoring of their performance as reflected in supervisor performance appraisals and in job-relevant training. The PATB was developed because the number of applicants for Agency jobs had increased to such an extent that it was not possible to perform intensive assessments on such a large number of people. The

development of PATB, which is a paper-and-pencil battery of tests that can be administered to a large number of people simultaneously, enabled PSS to obtain job-relevant information on a much larger number of applicants than had previously been possible through intensive assessment. By design, the test battery was developed to tap the same type of job-relevant information yielded by the PSS intensive assessment procedure, which in turn traces its origins back to the OSS days of World War II. Some of the job settings analyzed extensively by PSS psychologists, and on which test-score profiles were developed, are as follows:

Intelligence Analysts

Economic Analysts

General Operations Officers

CA Ops Officers

FI Ops Officers

CI Ops Officers

PM Ops Officers

Ops Support Officers

Log Officers

Personnel Officers

General Admin Officers

These are some of the job groups for whom PATB was originally developed and normed. The inspectors failed to mention these in their report. They also failed to mention the large number of job groups subsequently added to this list over the years the PATB has been in use. Part of the problem here is that the views of the IG consultants on job analysis are not up-to-date. The consultants, and consequently the inspectors who are guided by them, advocate a mechanistic, task-oriented approach to job analysis and recommend use of a device such as McCormick's PAQ (Position Analysis Questionnaire) to analyze Agency jobs. We in PSS do not recommend this approach because of the great diversity of professional jobs in the Agency. We favor instead job-analysis techniques based on determination of the attributes necessary to perform successfully in them, which is the way in which the content of PATB was originally decided upon in the first place. Today, most personnel psychologists would agree with us. In one of the most comprehensive and thorough

research reviews on job analysis written in modern times, Pearlman states in the January 1980 issue of Psychological Bulletin that "...the empirical evidence to date indicates that there is no necessity for molecular analyses of the specific tasks or elements of individual jobs and for complex grouping methods based on such analyses, at least when developing job families for purposes of validity generalization." Pearlman then proceeds to argue for the "...basic rationality, relative simplicity, and consequent generalizability of the human attributes underlying the structure of work, rather than the apparent chaos, complexity, and situational specificity of job analysis and test validation results that have long characterized the field."

Thus, the original developers of PATB were wise and foresighted in their approach to the analysis of Agency jobs and their subsequent utilization of this information in the selection of the PATB tests and subtests. We feel comfortable that we can refute the claims of the inspectors and

CPYRGHT

their consultants that the development of PATB was not based on a logical analysis of Agency jobs. Overwhelming evidence exists to the contrary.

- (5) On page 7 of the inspection report, the sweeping assertion is made that "There are no validity data of any kind...for the Strong-Campbell Interest Inventory..." This statement again misleads the reader, for it suggests that there is no value in administering this test. The main difficulty here is that the inspectors evidently have confused two types of validity. The Strong-Campbell Interest Inventory (SCII) and its predecessor, the Strong Vocational Interest Blank (SVIB), are the most widely used vocational interest tests in existence today and over the years have been normed and validated on tens of thousands of people representing a large number of occupational groups. The general value of these instruments is a matter of public record, as even the most cursory examination of the psychological literature readily shows. However, SCII scores do not predict performance on the job, and are not used for this purpose in

PSS reporting. Current description of Agency applicants in terms of vocational interest patterns is entirely proper and correct, and is fully consistent with the basic purpose for which the SCII and SVIB were initially developed. Further, we use the scores precisely as best practice, defined by the American Psychological Association, dictates.

- (6) The inspectors' report states on page 7 that there is no evidence for the validity of the writing sample. Either the inspectors do not understand how the writing sample (PATB Essay) is used or they have chosen to ignore the information provided them. We are informed from all sides that many jobs in the Agency require the ability to write well. To make estimates of an individual's use of grammar and syntax, not to mention spelling and punctuation, we ask for an actual writing sample. This is what the PATB Essay is. Managers who must select applicants with good writing ability apparently agree, and in many cases routinely request copies of applicants' PATB writing samples so they can make judgments of writing ability for themselves. The utility of a writing sample for such a purpose is intuitively obvious.

- (7) The inspectors also state on page 7 that evidence for the validity of the Biographical Information Inventory (BI) "...is lacking." The BI is used to identify applicants with unusual backgrounds for whom the PATB may not be fully appropriate. Low scores on cognitive ability measures, for example, may be caused by such factors as bi-lingual background, disadvantaged cultural and educational background, etc., all of which are ascertainable from examination of the PATB BI. The use of the BI for this purpose improves the interpretation of other PATB test scores. Used in this way the BI contributes substantially to the overall validity of the test battery. We thought the inspectors understood this.
- (8) We cannot understand the basis for the statement made on page 7 of their report to the effect that the PATB has "...little or no value..." for predicting success in foreign language training. Note: (a) PSS work in this area is of very recent origin and is at this point primarily exploratory in nature; (b) the number of languages studied to date is small because of the requirement for a sizeable data

base for each language; and (c) the results of this work are not used in making applicant selection decisions, which is the topic the inspectors were supposed to address. The inspectors' dismissal of this work as having "...little or no value..." does not square with the strong pleas of Agency language trainers to assign this research program as high a priority as resource availability will permit. The views of the inspectors are also at odds with the views of linguistics experts at the Center for Applied Linguistics, who have requested for external publication some of our findings in this area. We have good reason to be proud of our achievements in this very difficult area.

- (9) The inspectors appear on page 7 to object to the statistical confidence levels employed in PSS validation studies. As we have shown in paragraphs (1), (2), and (3) above, the validity coefficients obtained in our research studies are expected to be extremely conservative estimates of true validity because of restriction of range and the

criterion problem. Hence, the validity coefficients obtained in our research underestimate true validity, probably by a substantial margin. The selection of statistical confidence levels in our validation studies is designed to minimize the likelihood of what are known as Type II errors of statistical inference, which are discussed in all elementary textbooks on statistics and quantitative methods. We do not accept the inspectors' claims that our selection of statistical confidence levels is inappropriate, and we regard this as another attempt on their part to cast PATB in as negative a light as possible.

PATB Narrative Reporting

Perhaps nowhere in the inspection of PATB are errors resulting from the superficiality of the investigation more evident than in the discussion of PATB narrative reporting presented on pages 7 and 8 of the inspectors' report. The criticisms of PATB narrative reporting presented there have no factual basis. The inspectors and their consultants unfortunately neglected to examine the raw test scores on which the PATB reports judged by them to be deficient were based. The inspectors also made no effort to determine how

the PSS psychologists who write PATB reports interpret test-score profiles. Consequently, we cannot accept the inspectors' criticisms of PATB narrative reporting. Objective judgments of misinterpretation or inappropriate reporting of test scores simply cannot be made in the absence of direct, side-by-side comparison of the narrative reports with the test scores on which those reports were based. The inspectors and their consultants failed to do this. Their singling out of Strong-Campbell Interest Inventory (SCII) scores as being "...consistently misinterpreted..." is also unwarranted and unsubstantiated. The Chief of the PSS Assessment Branch, who is licensed as a professional psychologist in the Commonwealth of Virginia, oversees all SCII reporting. His credentials for SCII reporting include personal training by the major author and developer of the SCII. Neither the inspectors nor their consultants have any direct working knowledge of the SCII, which is a modern-day vocational interest inventory introduced in the early 1970's. The consultants have evidently confused the SCII with its predecessor, the Strong Vocational Interest Blank, which was originally developed by one of their contemporaries, Edward K. Strong, in the late 1920's. Perhaps most telling is the failure of the inspectors and their consultants to substantiate their criticisms with any evidence that PATB narrative reports have been misused. The best they can do is label PATB narrative reports as "...potentially unfair."

Relevance for Minorities and Females

The statement is made on page 9 of the inspection report that no "...studies of adverse impact as defined by EEOC have been done for PATB..." This astonishing statement is flatly contradicted by data made directly available to the inspectors and their consultants. In fact, on the very next page of their report, the inspectors cite a large-scale PSS study on possible adverse impact and conclude that the study is "...not conclusive with respect to fairness of the PATB." The inspectors go on to say "...that PATB does not appear to have any more bias than do...other selection procedures." These statements and conclusions are, we feel, deliberately misleading.

The study referred to attempted to determine whether there is adverse impact of PATB on the hiring of blacks for professional employment in CIA. There were three phases to the study: In Phase I, the selection ratio for 514 black applicants PATB tested between January 1974 and January 1977 as part of their applicant processing was compared with the selection ratio for 438 black applicants for the same time period who did not take PATB as part of their applicant processing. The selection ratio for the tested blacks was 15.0% hired. The selection ratio for the non-tested blacks was 14.1% hired. These selection ratios are statistically identical, which is one way to show absence of adverse impact. Application of the 80% rule, which the inspectors appear to

be unfamiliar with, yields a value of 94%, which further confirms the absence of adverse impact. PATB testing clearly does not decrease a black applicant's chances for employment with CIA.

In Phase II of the study, we compared the test scores of successful black applicants with the test scores of unsuccessful black applicants and found the scores of the first group to be superior overall to the scores of the second group. Hence, the Agency is tending to select those blacks most likely to succeed in Agency job settings, at least as reflected by PATB scores. Comparison of the PATB scores of a sample of successful and unsuccessful white applicants yielded virtually identical results. Thus, the PATB contributes to the selection for employment of both the best suited blacks and the best suited whites from their respective applicant pools. This is an important finding in that it shows that PATB is being used in essentially the same way to select blacks as to select whites.

In Phase III of our study, we compared the selection ratio for the 514 black applicants PATB tested between 1974 and 1977 with the selection ratio for [redacted] white applicants for the same time period who also took PATB as part of their applicant processing. Seventy-seven of the 514 PATB tested blacks entered on duty, yielding a selection ratio of 15.0%.

STATINTL

STATINTL [redacted] PATB tested whites entered on duty, yielding a selection ratio of 13.8%. Since the

selection ratio for the tested blacks exceeds the selection ratio for the tested whites, it is patently obvious that PATB testing has no adverse impact on blacks. Before leaving this, however, a check to see whether PATB testing has an adverse impact on whites is in order. Dividing the smaller selection ratio (13.8% for the tested whites) by the larger selection ratio (15.0% for the tested blacks) yields a quotient of 92%. Since 92% exceeds 80% by a comfortable margin, the PATB is acquitted of possible charges of adverse impact on whites.

In summary, the findings from this important study fully meet all requirements for determination of adverse impact as spelled out in the current EEOC Guidelines. Despite their efforts to malign the PATB on grounds of adverse impact, the inspectors are reluctantly forced to conclude on page 9 of their report that there is "...no direct evidence of bias or unfairness..." Apparently disappointed by their failure to find factual evidence of any kind that PATB is or has been misused, the inspectors cite on page 9 other factors which to them "...indicate a serious potential for misuse or unfair use of PATB..." Our response to these is as follows:

- (1) The inspectors claim that "...there is no evidence that minorities were represented in the groups used to norm the tests..." or in "...the samples used to determine

job-related validity..." They seem to imply here that a deliberate attempt has been made to exclude minorities from PATB studies. This claim on their part is absurd. Our policy in research studies is and always has been to use the largest sample size we can possibly obtain. The same policy applies to females as well. Minorities and females have always been represented in our research samples to the maximum extent of their availability. We reject the claims to the contrary made by the inspectors, who again offer no empirical data to substantiate their allegations.

- (2) The statement is made on page 9 that the equations being used to make recommendations for hiring decisions are "...based on inadequate samples of small size and have not been cross-validated." This statement is incorrect and misleading. The equations referred to here have for the most part been developed solely for research purposes and are not used in the general reporting of test results to consumers. An attempt is

made to imply that PSS is somehow responsible for the fact that some of this work is based on small samples, but the inspectors fail to mention that these "...inadequate samples of small size..." in many cases contain nearly 100% of the individuals employed in the particular Agency job category under study. The inspectors' claim of no cross-validation is inaccurate. In one instance we do report results from a statistical regression to a particular consumer in response to special request. These results are used by that consumer for a special purpose. The statistical regression technique used in this instance has not only been successfully cross-validated in terms of its ability to predict a certain type of highly relevant job behavior, but also has been demonstrated to possess a high degree of construct validity as well. Since the inspectors were provided a copy of a lengthy PSS research report describing this work in detail, their blanket condemnation of our research using statistical regression techniques is puzzling.

Summary Review

Before commenting on the inspectors' overall assessment and recommendations concerning PATB, it is appropriate to review briefly the conduct of the inspection and the findings as reported. We shall comment in the next section on the inspectors' specific recommendations.

- (1) Conduct of the inspection. In their approach to the inspection, the inspectors and their consultants made several implicit assumptions about PATB and PSS which are incorrect. They evaluated PATB and PSS (a) as if PATB is intended to be validated and used in a line fashion as a screening device for all employees, and (b) as if PSS has been given the resources needed for ongoing development and refinement of PATB, as well as line access to the performance appraisals needed to conduct validation research. Since these assumptions are incorrect, the inspection based on them was misdirected, and the general conclusions emanating from this effort were incorrect.
- (2) Selection of consultants. The selection of the outside consultants was unwise, and resulted in an unfortunate misalliance of (a) in-house inspectors without professional

training and expertise in the use of tests in personnel selection, and (b) outside consultants constrained by purely academic and out-of-date views on testing. As a result, the investigation was misguided and its conclusions and recommendations flawed.

- (3) Role of PATB in personnel selection. The inspectors failed in their stated objective to determine how PATB is used in personnel selection in the Agency. They did not, for example, determine (a) what inputs managers use in making personnel decisions, or (b) how PATB is weighted in relation to other inputs. This aspect of the investigation contains errors which call into question other aspects of the inspection effort. For example, the inspectors state on page 2, evidently in an effort to imply that PATB has little value, that "...the longer the time of service or supervisory experience..., the less the value placed on PATB's usefulness." This conclusion is based on a comparison of responses on items from the inspectors' survey administered to 500 Agency supervisors. The inspectors' conclusion requires significant correlations between the responses to Item 3 (How long have

you been employed by CIA?) and Item 4 (How long have you been a supervisor...?) with the responses to Item 13 (Which statement below most reflects your opinion of the usefulness of...PATB...?). However, the responses to Items 3 and 13 and to Items 4 and 13 are not statistically related. Hence, the inspectors' claim is based on incorrect interpretation of very simple and straightforward empirical data. The inability of the inspection team to interpret correctly their own data renders the credibility of other aspects of the inspection based on interpretation of survey responses highly suspect.

- (4) PATB reliability. The inspectors and their consultants attempted unsuccessfully to impugn PATB on grounds of inadequate reliability. Data presented by us (see Table 1) refute the efforts of the inspectors, and show, based on the seven PATB tests examined, that the reliability of the battery is excellent. Reliability is demonstrated for females as well as males, and for blacks as well as whites.

(5) PATB validity. Using an inappropriate model which works well in academic textbooks, but not in applied organizational settings, the inspectors argue incorrectly that evidence for the validity of PATB is unconvincing. Both the inspectors and their consultants failed to consider properly (a) the attenuation expected in validity coefficients when employees cannot be randomly selected, and (b) the pattern of validity coefficients expected when factorially simple tests are correlated with factorially complex job-performance measures. Statistically significant validity coefficients have been found in every major PSS validation study undertaken since the introduction of PATB. The validity coefficients obtained in these studies are conservative estimates of true validity; they are consistent with the factors ignored by the inspectors in their review of PATB validity as noted in points (a) and (b) above. The inspectors also disregarded substantial evidence available to them and erroneously concluded that the original selection of PATB tests and sub-tests was not based on logical analysis of Agency jobs.

- (6) PATB narrative reporting. The inspectors and their consultants present a strident denunciation of PATB narrative reporting. In no instance did the inspectors compare the narrative reports they fault so strongly with the actual test scores on which those reports were based. Hence, the inspectors' claims of faulty and inappropriate reporting have no factual basis whatever, and are at odds with the views of our own consultants, who find our narrative descriptions of applicants based on test scores conservative and well within the bounds of available data.
- (7) Relevance for minorities and females. The inspectors and their consultants make a concerted but unsuccessful effort to malign PATB on grounds of potential adverse impact on minorities. They also imply that minorities and females have been systematically excluded from studies of the test battery. These allegations are false, and are flatly contradicted by evidence directly available to the inspectors and their consultants. We are at a complete loss to explain their utter disregard of data collected by PSS

which show convincingly, and in the manner prescribed by the EEOC Guidelines, that PATB testing has no adverse impact on blacks.

In summary, we believe that the inspection of PATB is a failure. We do not believe that the inspectors and their consultants were sufficiently thorough or properly objective in their examination or in their reporting on PATB. Despite their unnecessarily negative reporting, which on occasion lapses into histrionics, the inspectors and their consultants are unable to cite a single instance where PATB is or has been misused. The language used by the inspectors to describe PATB includes such terms as "potentially unfair" and "serious potential for misuse or unfair use," but again we must stress the fact that the description of PATB in these terms is not based on factual evidence of any kind that misuse of PATB has actually occurred.

IG Assessment and Recommendations

We do not accept the recommendations made by the inspectors. Their recommendations are based on flawed reasoning and serious errors in interpretation of available evidence. Our views are shared by our consultants.

In Recommendation 1, the inspectors suggest that the use of several PATB sub-tests be discontinued. These are as follows:

- (1) The inspectors recommend discontinuation of the Strong-Campbell Interest Inventory.

This recommendation is without foundation and reflects bias toward this instrument evident from the outset of the inspection.

As far as we can determine, the views of the inspectors and consultants toward this instrument are based on misunderstanding of how it is used and confusion on their part between it and its predecessor, the Strong Vocational Interest Blank. We have been assured by no less an authority than the principal author and developer of this instrument, whose name appears on the title page of this report, that our use of his inventory is correct, conservative, and totally consistent with the purposes for which it was designed.

- (2) The inspectors recommend discontinuation of the PATB test of foreign language aptitude. This comes as a complete surprise considering the critical dependence of the CIA on foreign language aptitude and competence. The inspectors and their consultants have no factual basis to support such a recommendation. The PATB Foreign Language Aptitude test was taken from the Foreign Language Aptitude Test Battery developed

and validated by the U.S. Army. The reliability of this instrument is exceptionally high (see Table 1). To discontinue use of a test with credentials this impressive in an organization as critically dependent upon foreign language aptitude as the Agency would clearly be inappropriate.

- (3) The inspectors and their consultants recommend discontinuation of both the PATB Work Attitudes and the Thurstone Temperament Schedule. These two instruments provide a glimpse of the individual's personality and are the only PATB tests that do so. Since these instruments are basically personality measures, their scores are interpreted clinically by PSS psychologists and are particularly useful in our intensive assessment activities, which the inspectors did not review. The inspectors and their consultants lack the necessary professional background to evaluate properly the utility of these instruments; hence, their recommendations for discontinuation are without foundation.
- (4) We have already presented convincing arguments for use of the PATB Biographical Inventory

(BI), which the inspectors also recommend be discontinued. The inspectors failed to determine how the BI is used. This instrument is used primarily to determine whether other PATB scores are adversely affected by factors in the individual's background. The use of the BI in this way improves the accuracy of interpretation of other PATB scores and consequently augments the overall validity of the test battery.

- (5) The inspectors recommend that PSS prediction equations not be used in the evaluation of applicants for job assignments. PSS prediction equations are used for research purposes. They are not and never have been used in the evaluation of applicants for employment. The inspectors knew this.
- (6) We have already responded to the inspectors' arguments against PATB narrative reporting, which they recommend be discontinued. Throughout their report the inspectors repeatedly warn about the "serious potential for misuse or unfair use" of test results. They recommend later in their report that narrative reporting be abandoned and replaced by the reporting of raw test scores. We know of no

better way to promote unfair and inappropriate use of test results than to provide raw, un-evaluated test scores which could then be used in a cutoff, pass-fail mode by managers un-trained in test-score interpretation, but who nevertheless must make hiring decisions. PATB scores are not and should not be used in a pass-fail manner. Safeguards against this are provided by the use of narrative reporting, in which a comprehensive description of the individual can be presented based on integration of all available test results and in which no raw test scores are presented. The inspectors' recommendation to replace narrative reporting with a list of raw test scores is ill-advised. The net result of this would be to increase rather than decrease potential misuse of test results.

Based on the discussion presented here, the Director of Medical Services joins us in rejecting completely all aspects of Recommendation 1 made by the inspectors and their consultants.

Recommendation 2 is based on the incorrect assertion, which we have commented on extensively, that the original construction of PATB was not based on logical analysis of Agency jobs. This assertion is shown to be incorrect by information made available to the inspectors in the PSS Test

Data Book. Part of the problem here is that the inspectors, who are guided by their consultants, advocate a molecular, task-oriented approach to job analysis, while we in PSS favor the attribute-determination approach to job analysis on which the selection of the PATB tests and sub-tests was originally based. Two factors favor an attribute-determination approach over a task-oriented approach. First, the Agency contains a large number of highly disparate professional jobs for an organization of its size, with relatively small numbers of employees assigned to many of the professional job categories. Because of this, efforts to group jobs on the basis of common task requirements are less likely to be successful for purposes of validity generalization than efforts to group jobs on the basis of common attributes required to perform successfully in them. Second, and perhaps more importantly, the Agency has typically been interested in selecting employees with long-term career potential, i.e., the ability to perform successfully in a progressive series of job assignments of increasingly greater responsibility, culminating ultimately, for those most successful, in assignment to senior management positions. Tests overlapping factorially with the attributes viewed as common to a progressive series of increasingly responsible job assignments are intuitively superior to tests which merely predict performance in carrying out simple behavioral tasks in entry-level jobs. This summarizes in part some of the thinking

behind the original development of PATB, which was designed to provide an estimate of career potential as well as performance in entry-level job assignments. The inspectors and their consultants seem to have missed this point completely, the consultants perhaps because of their narrow textbook approach to testing and job analysis, the inspectors perhaps because of their lack of expertise in these areas.

We turn now to Recommendation 2.a., which appears on page 13 of the inspectors' report. We reject this recommendation as stated for several reasons. First, we reject the conclusion that McCormick's Position Analysis Questionnaire (PAQ) is the most appropriate instrument for analysis of Agency jobs. We do agree that more job analysis is needed, but we favor an attribute-determination model for the reasons outlined above. Second, we do not support the recommendation that the responsibility for job analysis be assigned to the Office of Personnel Policy, Planning, and Management (OPPPM). The inspectors apparently do not understand the difference between a job analysis, which belongs within the domain of employment test validation, and a job description, which belongs within the domain of position management and compensation, and is, of course, a legitimate OPPPM function. The difference between job analysis and job description is considerable. If, in the judgment of responsible Agency officials, more job analysis is needed, then the

responsibility and authority for this should be assigned to OMS and to PSS, where the primary charter to provide psychological testing support to the Agency currently resides.

We also reject completely Recommendation 2.b. appearing on page 14 of the inspection report that a selection test development unit be established under the Director of OPPM. Again, if responsible Agency officials conclude that additional work on PATB is called for, then the responsibility for this work, and the personnel resources needed to carry it out, should be assigned to PSS.

Before leaving Recommendations 2.a. and 2.b., which in essence call for acquisition of test development specialists and personnel psychologists with training in job analysis, we should examine the current availability of these resources within the existing PSS resource package. A quick look at our overworked research unit reveals a shortage of personnel needed for ongoing refinement, improvement, and validation of PATB for new job settings. One can legitimately ask why, when the PATB was initially developed, sufficient personnel resources were not set aside for ongoing maintenance and further development of the battery. The fact is that they were, and research on PATB flourished during the first few years the battery was in use. The principal reason for the sudden diminution in research on PATB can be found embedded within the institutional memory of the IG staff itself. An IG investigation of PSS conducted in 1959 seriously questioned

the continuing need for an ongoing in-house research program on the test battery, and recommended a reduction in the size of the PSS research unit, which followed as a result. In 1968, however, the IG staff gave us high marks for our research in support of PATB testing. They concluded in their 1968 inspection that PSS "...deserves a great deal of credit for developing, refining, and validating its screening tests for professionals. This essential task has been performed in a most competent manner; its value is recognized throughout the Agency." Now in 1980, the IG staff reverses its conclusions of 1968 and at the same time recommends reinstatement, although under a different organization, of the in-house personnel and programs it recommended be eliminated in 1959! In view of these inconsistencies, we can hardly see, methodological inadequacies notwithstanding, how the results of the 1980 IG inspection of PATB can possibly be taken seriously.

We also reject Recommendation 2.c., which calls for the establishment of centralized policies as to who should take tests and who should not. We favor retention of current Agency practice under which individual managers decide for themselves when to use PATB as an adjunct to the selection process.

The Director of Medical Services joins us in rejecting Recommendations 3, 4, and 5, all of which would make raw, unevaluated test scores available to individuals not trained

in their interpretation and proper use. As we pointed out earlier, substitution of raw, unevaluated test scores for narrative test reports increases rather than decreases the potential for misuse of test results.